THERMODYNAMICS AND STRUCTURAL CHARACTERIZATION OF BAR DOMAIN DIMERIZATION FROM MD SIMULATIONS

by
Adip Jhaveri

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of

Master of Science in Engineering

Baltimore, Maryland
May, 2020

Abstract

Protein-protein interactions are essential steps in nearly all biological processes. The BAR domain proteins are a large family that interact with one another in solution and on the membrane to help drive membrane remodeling. Here the binding thermodynamics of homodimerization between the Lsp1 BAR domain proteins in solution, is studied using MD simulations. By combining coarse-grained protein models with enhanced sampling through metadynamics, we are able to construct a free-energy surface describing the bound vs unbound states along multiple collective variables. From these surfaces, the K_D values are computed as well as the relative entropic and enthalpic contributions. In addition, the results are verified to be robust under variations to the parameter selections in the metadynamics approach. The stuctural intermediates encountered during the binding process are also characterized. With these results, a rich and quantitative perspective on the binding thermodynamics of moderately strong protein-protein interactions is provided, that is representative of a wide range of protein contacts that are critical for cell biology.

Primary Reader and Advisor: Dr. Margaret Johnson

Secondary Reader: Dr. Paulette Clancy

Acknowledgements

I am grateful to my advisor, Dr. Margaret Johnson for giving me the opportunity to work on this project and for her support during the last 18 months. Dr. Johnson’s clarity of thought and lucid explanation of her ideas have been instrumental in helping me improve my work. I look forward to working with her for my doctoral research.

I’d also like to thank my lab members, past and current. Matt for introducing me to the tools essential to the project and helping me with all sorts of problems I had faced during my initial days. Yiben, for our discussions and providing timely inputs to my work and in general for being a great role model. Bhavya, also providing additional suggestions over the course of the last year regarding my work and presentation. Dhruw, for providing me the results of his simulations.

And lastly, a shoutout to all the lab members, I may have missed out, for being a great group of people I enjoy to working with.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 BAR proteins

The phospholipid bilayer present in cells, while acting as a barrier to the external environment, is extremely dynamic in nature. The flexible and dynamic nature of the bilayer allows the cell to actively interact with its environment. Specifically, membrane remodeling underlies essential cellular pathways of endocytosis, motility, signaling and division. Beneath this complex process is a network of protein-protein interactions that carry out specific tasks with high spatial and temporal accuracy. An important type of these proteins are the BAR domain family of proteins that induce curvature upon binding to membrane.

The Bin-Amphiphysin-Rvs (BAR) domain protein were first recognized for possessing evolutionary conserved regions and were implicated to have a functional role in cell division, actin dynamics and membrane trafficking [1,2]. With the identification of crystal structures of Arfaptin and amphiphysin, BAR proteins came to be known for their characteristic crescent shaped dimer form [3,4]. The BAR domain superfamily of proteins consists of the classical BAR, N-BAR(N-terminal amphiphatic helix), F-BAR (Fer/CIP4 homology-BAR) and I-BAR (IRSp53 MIM homology-BAR) and are known to be involved in membrane remodeling through different

mechanisms [5]. Experimental studies using crystallography and electron microscopy have revealed insightful structural details that have improved our understanding about this process. The BAR dimer is stabilized by hydrophobic interactions between the inner surface of the monomers. A common feature among the BAR proteins, is the concentration of positively charged residues at the concave surface which predominantly forms the membrane binding interface [1,5]. As the BAR proteins oligomerize on the membrane surface, the curvature is primarily driven by electrostatic interactions and the intrinsic shape of the BAR dimer. Another possible mechanism is observed for BAR proteins exhibiting an additional amphiphatic helix. This involves insertion of the helix, which acts as a wedge, and rearranges lipids due to hydrophobic interactions and causes the membrane to bend [2,5].

1.2 Molecular Dynamics

Molecular dynamics is a simulation method to generate configurations of a system by integrating Newton's laws of motion. This results in the evolution of atomic positions and velocities through time, providing information on their collective behavior. MD simulations were first used to study fluid systems offering valuable insight into their structural and dynamical properties [6,7]. For protein molecules, it was first used to investigate the dynamic motions which are not captured by experimental techniques [8]. Since then MD simulations have formed a bridge between theoretical and experimental investigation and have been used to estimate thermodynamic and transport properties of complex systems that cannot be calculated analytically.

The simulation of molecules, in time, can be achieved by solving the equations of motions based on first-principles physics. This would concern evaluating the motions of electrons and nuclei

which are described by the Schrodinger wave equation [9]. However, for biological molecules, this equation is infeasible to solve computationally. An alternative approach is molecular mechanics, which follows classical mechanics to evaluate the forces acting on atoms. The forces are modeled based on an underlying force-field that describes potential energy function of atomic nuclei while regarding electronic motion implicitly (Born-Oppenheimer approximation) [10].

The force-field represents a model of inter and intra molecular forces by simple energy functions of the internal coordinates of the system. The force-fields are empirical and designed based on parameter fitting to quantum mechanical calculations and experimental data to reproduce structural and thermodynamic properties of molecular systems. An example of a functional form is shown below [10]-

$$U = \sum_{bonds} \frac{1}{2} k_b (r - r_0)^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{torsions} \frac{V_n}{2} [1 + \cos(n\phi - \delta)] + \sum_{i,j} 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) + \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0 \epsilon_r r_{ij}}$$

Here the potential energy is described as a sum of individual energy terms arising from different types of interactions. The first term represents the energy between bonded atom pairs, modeled as a harmonic potential where deviations from equilibrium of reference bond length are penalized. The second term represents the energy due to angle bending, modeled as a harmonic potential function of the angle between two adjacent bonds. The third term models the energy due to rotation of bonds, called a torsion potential. The fourth term represents all the non-bonded interactions. The pairwise electrostatic interactions are modeled by a Coulomb potential and the van der Waals interactions are modeled by a Lennard-Jones potential. Majority of the force-fields

naturally contain these four components (albeit in different functional forms) with additional sophisticated terms to attain higher accuracy. However, the price usually associated with obtaining higher accuracy is computational efficiency. The choice of a force-field depends on the system of interest, goal of the study and the parametrisation process.

After the selection of a suitable energy model, the simulation of the system proceeds in four different stages: i) Solvation ii) Minimization iii) Equilibration iv) Production phase.

MD simulations offer a distinctive advantage of sampling configurational space and observing the dynamics of a system at increased spatial and temporal resolution. As a result simulations have been applied to reduce ambiguity in force microscopy measurements [11], resolve structural data in electron microscopy maps [12] and often predicting unique insights. In addition, MD simulations have been increasingly used towards evaluating the thermodynamic properties of systems.

Evaluating thermodynamics from MD simulations

The observed properties of a system that define a macroscopic state arise from the microscopic dynamics or fluctuations of all the particles in the system. To obtain an estimate of macroscopic properties would involve computing all the variables that define the microscopic state of each particle [13]. For a classical system, this is equivalent to the phase space (r^N, p^N) where r^N corresponds to the coordinates of N particles i.e. (r_1,r_2,…,r_N) and p^N corresponds to the conjugate momenta i.e. (p_1,p_2,...,p_N). However, evaluating the dynamics of N particles for a macroscopic system ($N \sim 10^{23}$) is infeasible. Instead we apply the principles of statistical mechanics that

relates macroscopic variables to probability distributions of microscopic states. An underlying assumption of statistical mechanics is that the measured value of a property corresponds to the average of that property over all microscopic states, i.e an ensemble average [13]. These microscopic states belong to a particular macroscopic state constrained by a small number of variables which could be for example the total number of particles, N, volume, V and temperature, T.

If we consider a classical system that is initially at a particular microscopic state and is then observed to evolve through time and visit different phase space points for sufficiently long time, then the time average of a property can be said to be equivalent to the ensemble average. This is called the principle of ergodicity [13]. MD simulations which propagate a system in time, therefore, are a powerful tool in calculating equilibrium thermodynamics quantities by sampling the configurational space. Certain thermodynamic properties are simple enough to calculate from simulations like internal energy (U), temperature (T) and heat capacity. Other properties like free energy and entropy which are usually of interest are difficult to obtain from simulations. An important quantity to introduce here is the partition function, Z, which is equal to the sum of the Boltzmann weighted Hamiltonian, of all possible microstates for a particular macroscopic state. If we consider the canonical ensemble, which is the assembly of all possible microscopic states with constant NVT, the equilibrium probability of observing a particular microscopic state (r^N,p^N) obeys -

$$P(r^N, p^N) \propto \exp(-\beta H(r^N, p^N))$$

Here H is the Hamiltonian of the system and β is the inverse temperature ($\beta = -1/k_BT$) and k_B is the Boltzmann constant. The constant of proportionality is obtained by following the normalization requirement $\sum P(r^N,p^N) = 1$. This leads to the expression [10]

$$P(r^N, p^N) = \exp\frac{(-\beta H(r^N, p^N))}{Z} \quad \textit{where} \quad Z = \frac{1}{N!h^{3N}} \int\int \exp(-\beta H(r^N, p^N)) dr^N dp^N$$

where Z(N,V,T) is the canonical partition function. The partition function conflates statistical mechanics with classical thermodynamics in a sense that macroscopic thermodynamic quantities can be calculated from the partition function. The Helmholtz free energy for example can be obtained by the following expression – $A = -k_B T \ln Z$

So to obtain an estimate of the free energy from MD simulations, the partition function needs to be resolved by sampling all possible microscopic states in the NVT ensemble. For macromolecules the timescales required to achieve this would demand an impractical amount of computational resources [14]. In addition large molecules like proteins possess a rough energy landscape which usually leads the system to be kinetically trapped in local energy minima regions and restrict the sampling of other important microscopic states. To overcome this difficulty, several algorithms have been developed that can interface with the MD code and improve the sampling of important states in phase space. Some examples of these are umbrella sampling, weighted ensemble and metadyamics.

A certain limitation in performing these simulations is accessing longer time-scales at which most of the most biomolecular events such as protein folding, protein-protein binding take place.

In addition, simulation of large systems involving millions of atoms interacting at meso-length scales are computationally expensive [9,14,15].

1.3 Metadynamics

Metadynamics is an enhanced sampling technique, that probes the system under a reduced set of degrees of freedom called collective variables (CVs) [16]. This involves adding an history dependent potential term (called a bias potential), as a function of the CVs, to the Hamiltonian of the system. The CVs, are a function of the atomic coordinates and can sufficiently describe all the states of the system. The bias potential (V(s)) at time (t) is of the form [17]-

$$V(s,t) = \sum_{t'=\tau_G, 2\tau_G \ldots}^{t} w \exp\left[\sum_{i=1}^{N_{CV}} \frac{\left(s_i(q) - s_i(q(t'))\right)^2}{2\sigma_i^2} \right]$$

During the simulation, the bias potential are deposited in the form of Gaussians centered on the explored points. Here τ_G is the time interval for adding the bias potential, w is the height of the Gaussian hill and σ is the width for each CV. The system now evolves according to the probability distribution $P'(s)$as given below -

$$P'(s) \propto \exp\left(- \frac{F(s) + V(s)}{k_B T} \right) \qquad P(s) \propto \exp\left(- \frac{F(s)}{k_B T} \right)$$

Here P(s) is the unbiased distribution. The free energy F(s) is defined as [18]-

$$F(s) = -k_B T \ln \int \delta(s - s(R)) \exp(-\beta U(R)) dR^N \quad + \quad C'$$

Generally, in an unbiased simulation, the system gets trapped in lower energy regions, it remains there for a long time, since the thermal fluctuations (of order ~k_BT) are not enough to overcome an energetic barrier [19]. However, in metadynamics, when the system is stuck in a local minima,

the bias potential accumulates as a sum of Gaussian hills in the CV space effectively flattening the free energy surface in that region. This drives the system to visit states that are separated by high energy barriers, states it would not have visited in an unbiased simulation. After sufficient time(t), the unbiased estimate of free energy can be recovered from the biased potential by-

$$V(s, t \to \infty) = -F(s) + C$$

Chapter 2

Objectives

2.1 Motivation

BAR proteins, as mentioned in Chapter 1, have been a focus of broad structural and functional characterization due to their central role in membrane remodeling. Despite the identification of common features across distinct sub-classes of BAR domain proteins, they each exhibit unique features that impart function and cannot be ascertained without additional study. The dimerization of BAR proteins is a crucial step in mediating protein-lipid and protein-protein interactions during several cellular pathways like clathrin mediated endocytosis [2,20]. The weak sequence homology across different BAR domains across indicates that each BAR dimer could be stabilized by different intermolecular reactions. For example certain amino acids at the BAR domain interface of the SNX33 (sorting nexin 33), have been identified to impart specificity for dimer interactions [21]. For endophilin A1, it was reported that a mutation of a single hydrophobic residue, located at the BAR domain interface, can impair dimerization [22]. In addition to identification of unique interactions that stabilize the dimer, this process requires further investigation in terms of the underlying thermodynamics. Experimental studies focused on characterizing the kinetics and thermodynamics of this process, have reported dimerization

affinities, in solution, in the range of subnanomolar to micromolar leading to different mechanistic explanations [22–24]. This demonstrates the complex nature of the different domains present in the proteins and the challenges to obtain a quantitative understanding of this process. Here the protein of interest is the yeast BAR domain protein Lsp1, known to be involves in eisosome formation [25].

Molecular Dynamics simulations are a powerful alternative to study the dimerization of BAR domain protein and obtain structural and thermodynamic information simultaneously. Past MD simulations of BAR domain proteins have only focused on characterizing the structural aspects of membrane deformation or interaction with other proteins [26–28]. Atomistic scale MD simulations have been crucial in revealing experimentally unresolved molecular details regarding curvature inducing mechanisms for N-BAR and F-BAR proteins [29,30]. For this study we have used coarse-grained (CG) simulations by performing the MD simulations under the MARTINI force field [31]. CG simulations would provide a significant speed up with respect to computational time due to the reduced degrees of freedom.

Various methods have been employed to calculate the free energy of binding from computer simulations, like Free Energy Perturbation [32], Thermodynamic Integration [33] and Potential of Mean Force [34]. Computations of free energy from MD simulations are associated with numerous challenges. A few of them are (i) Choice of a suitable force-field (ii) sampling of all relevant states (iii) treatment of long-range interactions and (iv) convergence [35,36]. Some of these challenges can be addressed, for example, by combining with additional protocols like Replica Exchange [37] and Umbrella Sampling [38]. The choice of an appropriate method depends not only on

the system under consideration but also on the aim of the study. For our system of interacting BAR domain proteins, in addition to obtaining an accurate estimate of the free energy of binding, it is also important to characterize the intermediate/metastable states. Therefore it is required that the simulation is able to cross energetic barriers and sample all possible conformational states. However the timescales required to achieve this is not feasible within the current computational means. Recently, large-scale simulations applying a weighted ensemble protocol have been able to predict binding pathways and calculate the binding rate constant for the barnase-barnstar complex [39].

To improve the sampling within a feasible timescale, we perform coarse-grained MD simulations coupled with the enhanced sampling technique of metadynamics [16]. A primary reason behind using metadynamics is because a protein-protein interface is much larger than a protein-ligand binding pocket and an accurate description of the binding process requires the consideration of multiple reaction coordinates or collective variables. Metadynamics offers several advantages compared to other methods in that it is not only able to efficiently reconstruct a multidimensional free energy surface (FES) but also identify novel intermediate states and transition pathways [40,41]. Different variations of this protocol have been developed to address certain challenges like convergence and choice of collective variable (CVs) which include Well-Tempered Metadynamics [42], Parallel Tempered Metadynamics [43] and Multiple Walkers Metadynamics [44]. In this study, we use the standard version of metadynamics to evaluate the FES of the homodimerization and to obtain the binding affinity (K_D).

2.2 Aims of the study

For this study, I had focused on completing the following major aims -

i) Compute the binding affinity (K_D) for the homodimerization of Lsp1 protein.

ii) Evaluate the thermodynamics of the dimerization process (ΔG, ΔS, ΔH)

iii) Characterize intermediates and binding pathways

Chapter 3

Methods

3.1 System setup

The Lsp1 structure was obtained from PDB under accession code 3PLT [45]. To study the dimerization process, chains A and B were used. The simulations were performed under two different force fields: a) MARTINI and b) AWSEM . For MARTINI, the protein structure was coarse-grained using the ELNEDIN model [46], which is a combination of an Elastic Network model (EN) with the MARTINI force field [47]. This model integrates a structure-based molecular description with a physics-based force field to study intermolecular interactions while maintaining the structural integrity of individual monomers . The CG structure was generated using the martinize.py script using the elnedyn forcefield option with the spring force constant as 500 kJ mol^{-1} nm^{-2} and cut-off as 0.9 nm.

3.2 Simulation details

The MD simulations under the MARTINI force field were performed using the GROMACS package [48]. The software package LAMMPS was used to simulate the protein system under the AWSEM force field. The simulation details are given below for each force field-

MARTINI

The protein complex was solvated using the standard MARTINI water model and neutralized using Sodium ions (NA+) in a simulation box of dimensions 24 nm x 24 nm x 24 nm. The solvated protein with the counter ions was minimized using the steepest descent algorithm. The system was then equilibrated for a total of 10 ns. First, the system was equilibrated under constant NVT conditions for 1ns using a timestep of 10 fs and then for 4 ns with a 20 fs timestep. The temperature was kept constant at 310 K using the velocity rescaling method with a time constant of 1 ps. This was followed by simulations under constant NPT for 1ns using a timestep of 10 fs and then for 4 ns with a 20 fs timestep. The pressure was kept constant at 1 bar using Berendsen coupling with a time constant of 12 ps. For the production runs, performed under NPT, the pressure coupling was shifted to Parrinello-Rahman while maintaining the integration time step of 20 fs. The bonds were constrained by the LINCS algorithm for all the simulations.

The simulations were performed under periodic boundary conditions and using the Verlet scheme to build the neighbour lists. The non-bonded interactions were cut-off at 1.2 nm. The coulombic interactions were modeled using the Reaction field with $\epsilon_r = 15$ and $\epsilon_{rf} = \infty$ to account for electrostatic screening. The Lennard-Jones (LJ) potential was smoothly shifted to zero within the cut-off distance.

Polarizable Water

One of the limitations of the standard MARTINI water model is that due to the absence of charges it behaves as a simple LJ fluid without the ability to screen electrostatic interactions [49,50].

While this can be approximated by assuming a relative dielectric constant, the polarization and screening effects should be modeled explicitly at interfaces between polar and non-polar phases. In our preliminary simulations, we found that the inter-molecular interactions are sensitive to the electrostatics. Therefore, we decided to use the polarizable water model to explicitly account for these effects and possibly improve the accuracy of the calculated binding affinity.

The water model used here is the refined polarizable water model (refPOL) [51]. The protein structure with Sodium ions was extracted at the end of the NPT equilibration run with the standard water model. The system was then solvated with the new refPOL waters using the GROMACS *solvate* method. This was followed by minimization using the steepest descent algorithm. The system was then equilibrated under a constant NVT condition for 5 ns with a time step of 10 fs. This was followed by constant NPT simulations, first performed with a 10 fs time step for 4 ns and then with a time step of 15 fs for another 4 ns. The same coupling methods were used to maintain constant temperature and pressure conditions.

The refPOL model works best when used with the Particle Mesh Ewald method for electrostatic interactions. The PME potential was smoothly shifted to zero at the cut-off of 1.1 nm. The relative dielectric constant was used with the recommended value of 2.5. The LJ interaction was also smoothly shifted to zero at 1.1 nm.

3.3 Metadynamics parameters

The selection of appropriate CVs for any metadynamics simulation has to take into account that they are not only good descriptors of the system but also a sufficient biasing coordinate. Since

we did not have a priori estimate of the FES, different sets of CVs were chosen which would allow to system to overcome barriers in case of hidden CVs. In addition, metadynamics in principle would be insensitive to the choice of CVs and would allow to assess convergence in the simulations.

The first set of CVs, included distances between the two protein chains. This CVs set is named as dist_CV. As shown in Fig (), d1 and d2 represent the distance between the two binding sites on chains A and B. The width of the Gaussians to be deposited was 0.1 nm for both d1 and d2. The Gaussian width was determined based on the fluctuations of the CVs, evaluated in a preliminary unbiased simulation. The height of the Gaussian used was 10 kJ/mol at the start of the simulation, but was dropped to 5 kJ/mol after sometime. This was after the bias had been judged to have filled the initial well and the system had escaped the bound state. Finally two different simulations were run with deposition times of 140 ps and 500 ps.

The second CVs set, called angle_CV, is represented in Fig(). Here the dist is the distance between a third site on the two protein chains. The variable ang is defined as the angle between two vectors joining the N to C terminal on each chain. The width used for the Gaussian is 0.046 for ang and 0.1 nm for dist. Two different simulations were performed with different Gaussian height and deposition time. One simulation was run with a height of 5 kJ/mol and deposition time of 140 ps, and a second one with a height of 3kJ/mol and deposition time of 500 ps.

The metadynamics algorithm was implemented using the plumed plugin [52].

3.4 Analysis and Visualization

For post MD analysis, the position coordinates were recorded every 200 ps for each simulation. All analysis were performed in python[53] with the software package MDAnalysis [54,55]. Clustering analysis was performed using the scikit-learn tool implemented in python [56].

For visualization, the coarse grained structures were first convert to an all-atom representation using the pulchra algorithm [57]. These structures were then visualized using the VMD tool [58–60].

3.5 Theoretical framework

Consider a dimerization reaction -

$$P + P \rightleftharpoons P_2$$

At equilibrium, the free energy difference can be written as -[61]

$$\Delta G^{o}_{unbound-bound} = \mu^{o}_{P} + \mu^{o}_{P} - \mu^{o}_{P2} = -RT \ln \frac{\gamma_P \gamma_P}{\gamma_{P2}} \frac{[P][P]}{[P_2]C^o} = -RT \ln \frac{K_D}{C^o}$$

Here μ_i^0 is the standard chemical potential of i, γ_i is the activity coefficient, which at low conc. can be assumed to be 1, C^o is the standard conc. (~ 1M) and K_D is the binding affinity. To compute the binding affinity, the standard state change in free energy needs to be evaluated from the simulations. Using metadynamics we obtain the following expression -

$$\Delta A_{meta} = A_{unbound} - A_{bound} = -k_B T \log \frac{Z_{unbound}}{Z_{bound}} \qquad Z_{bound} = \frac{\int_{bound} e^{\frac{-F(s)}{k_B T}} ds}{Q} \qquad Z_{unbound} = \frac{\int_{unbound} e^{\frac{-F(s)}{k_B T}} ds}{Q}$$

Here Z_{bound} and $Z_{unbound}$ represents all configurations in the bound and the unbound states respectively. Q is a normalization factor that is same for both the states. To obtain a meaningful comparison with experimentally calculated quantities which are in reference to a standard state, we have to add a correction term [62] and account for the work done due to fluctuations in volume under constant pressure conditions. The final expression is given below -

$$\Delta G^{o}_{unbound-bound} = \Delta A_{meta} - RT \ln[\frac{V_P V_P C^o}{V_{P2}}] + P^o \Delta V_{P2}$$

Here V_P and V_{P2} is the volume of the system in the bound and unbound state respectively. From this expression, the change in entropy can be calculated by the relation [63]-

$$\Delta S^{o}_{unbound-bound} = -\left(\frac{\partial \Delta G^{o}_{unbound-bound}}{\partial T}\right)_P$$

which gives the following expression-

$$\Delta S^{o}_{unbound-bound} = -\frac{1}{T}(\Delta G^{0}_{unbound-bound} - P^o \Delta V_{P2}) + \frac{1}{T}\left(\langle U \rangle_{unbound} - \langle U \rangle_{bound}\right)$$

Chapter 4

Results and Discussion

4.1 Free Energy Surface (FES)

The FES, of the simulation dist_CV, was evaluated after a total time of 88 μs. The free energy as a function of the CVs, d1 and d2, is displayed as a contour plot in Fig. 4.1(b). Under the influence of the bias potential, the protein chains visit different configurations in the CV space, starting from an initial bound structure. The crystal structure is represented by d1=0.8525 nm, d2=0.7766 nm (depicted by the ‘X’ point on FES). As observed from the FES, the metadynamics protocol is able to identify meta-stable configurations of the BAR dimer. As the system spends more time in these local minima regions, sampling different microscopic configurations for that region of CV space, the Gaussian hills accumulate until the system is able to cross the energy barrier and explore additional regions. The time evolution of d1 and d2 is shown in Fig. 4.1(c). The CVs d1 and d2 are able to identify the different states of the system including a vast volume of the unbound region. Over an 88 μs run time, we have observed 4 transitions, where the two chains, from an unbound state, were able to find a bound state close to the crystal structure configuration. The transition events have been pointed out in Fig. 4.1(c).

With an estimate of the free energy obtained from the metadynamics protocol, I proceeded with calculating the free energy difference between the bound and the unbound states. However, an important criteria to consider while evaluating the thermodynamics in this case is defining the bound and the unbound states. In addition, one cannot simply take the minimum point of the energy well of that state. Since the CV is a low-dimensional space possessing degeneracy

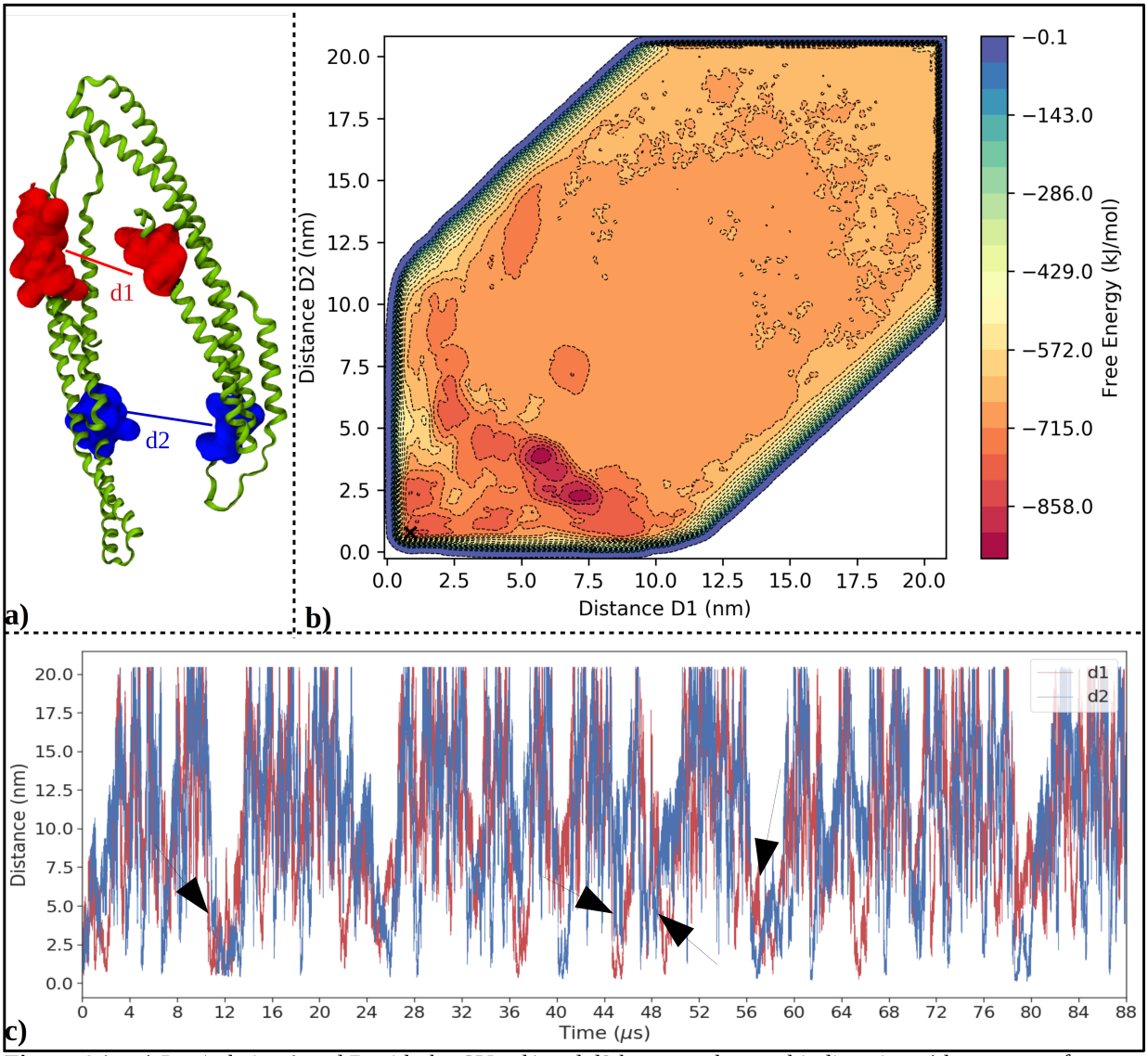


Figure 4.1 – a) Lsp1 chains A and B with the CVs, d1 and d2 between the two binding sites. (shown as surface representation in red and blue). b) Contour plot of the free energy as a function of d1 and d2 as calculated from metadynamics after 88 μs. The '**X**' represents the CV values for the crystal structure. c) Progression of d1 (red) and d2 (blue) with time. The arrows represent transition events.

with respect to the microscopic configurations, it is important to take into account the fluctuations of the CVs in each state. To define the bound states, I have used the following criteria:

- Choose the crystal structure and regions around it in the CV space as bound state
- Apply clustering on d1 and d2 to classify different regions of meta-stable configurations
- Use the RMSD w.r.t crystal structure to define bound state

Clustering

Clustering is a useful method to classify all sampled configurations and identify similar structures depending upon a particular set of features. Here, I performed KMeans clustering[64,65] using d1,d2 as features for n=6 clusters. The clustering was performed for only those configurations where chain A and B are in contact with each other. A contact was defined when the distance between any residue pair of chain A and B is less than 5 A^{o} . This trajectory consisted of only these configurations connected by time, will be called the bound trajectory. The generated clusters are shown in Fig. 4.2(a). In addition, the Cluster C-4, was subjected to another round of KMeans with same features and n=4, as shown in Fig 4.2(a). This sub-cluster provides an improved resolution to identify the cluster closest the bound state, which here would be SC-1 (olive green in Fig. 4.2(b)). The bound state was then defined by considering the CV bin and its nearby regions, of the closest configuration to the centroid of the cluster SC-1 .

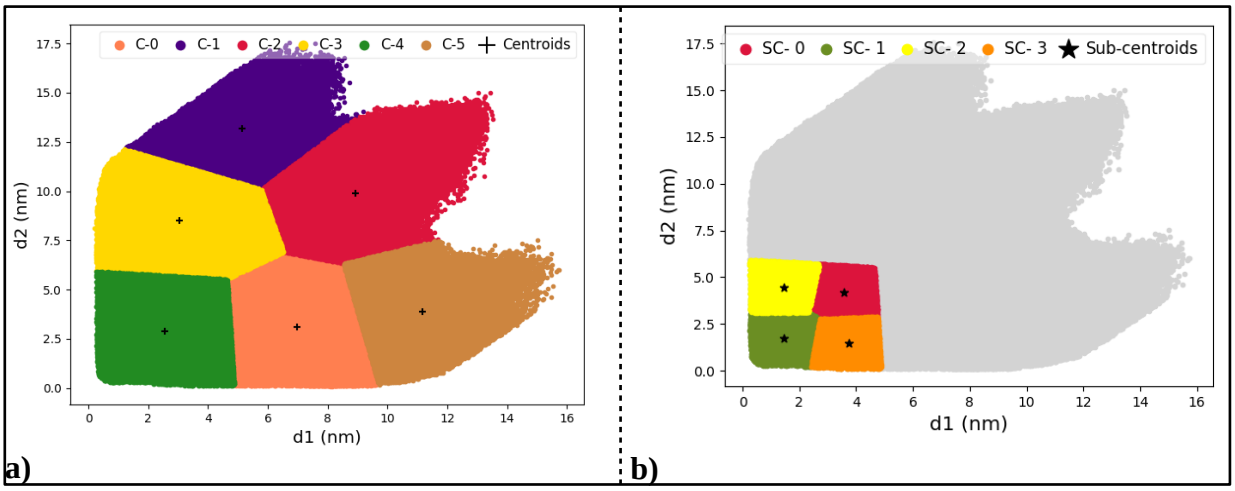


Figure 4.2 – a) Plot of d1,d2 colored by each cluster. b) Sub-cluster representation for C-4.

The free energy for the two bound states defined by using criteria a) and b) was calculated. These regions are highlighted in yellow on the FES in Fig. 4.3. In addition to the free energy, the average internal energy and average volume of system were calculated and are shown in Table-1.

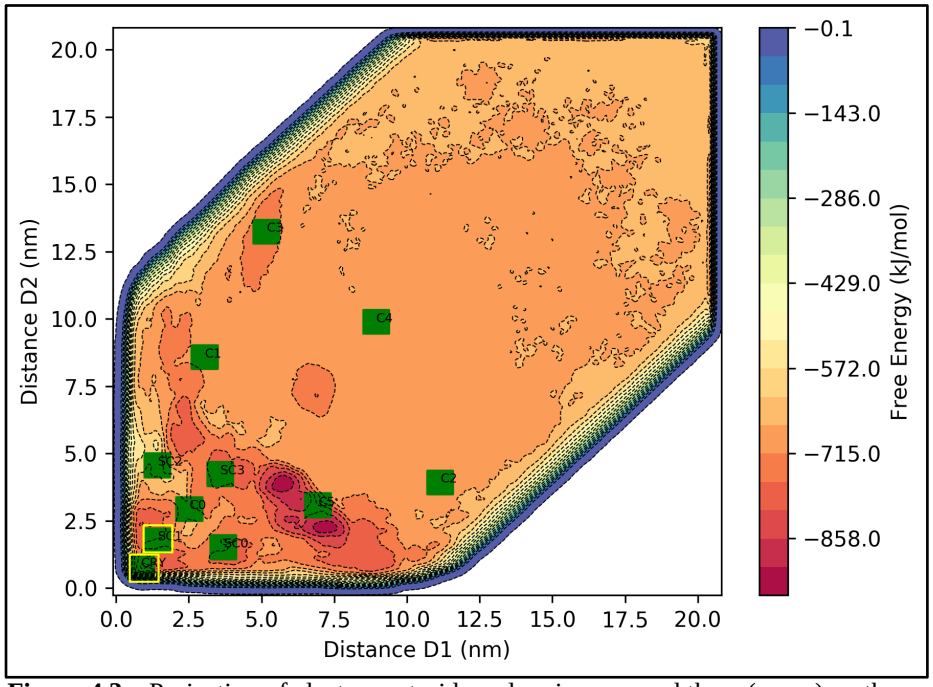


Figure 4.3 – Projection of cluster centroids and regions around them (green) on the FES. The two bound state regions are highlighted in yellow.

For criteria c), the RMSD was calculated according to the following steps- 1) Each configuration was was superimposed on the crystal structure w.r.t. chain A. 2) The RMSD was then calculated for the distances of each pair of binding sites on chain A and B of each configuration w.r.t the crystal structure positions. The RMSD was calculated for the bound trajectory (as defined above) and is plotted in Fig. 4.4(a). Since the superimposing is performed with chain A, the chain B RMSD is able to define different states of the system including transition events. For defining the bound state, only those configurations were chosen whose chain B RMSD was less than 1 nm. The free energy was calculated by integrating the CV bins corresponding to the selected configurations based on RMSD which is highlighted in Fig. 4.4(b). The free energy with the average potential energy and average volume are shown in Table-1.

Table-1 Thermodynamic quantities for the defined bound states

State	**Free Energy (kJ/mol) (A_{bound})**	**Potential Energy (kJ/mol) (U_{bound})**	**Volume (nm3)**
Crystal Structure	-814.18	-3.177 e^6	13140
SC1	-776.28	-3.177 e^6	13141
RMSD based	-822.032	-3.177 e^6	13141

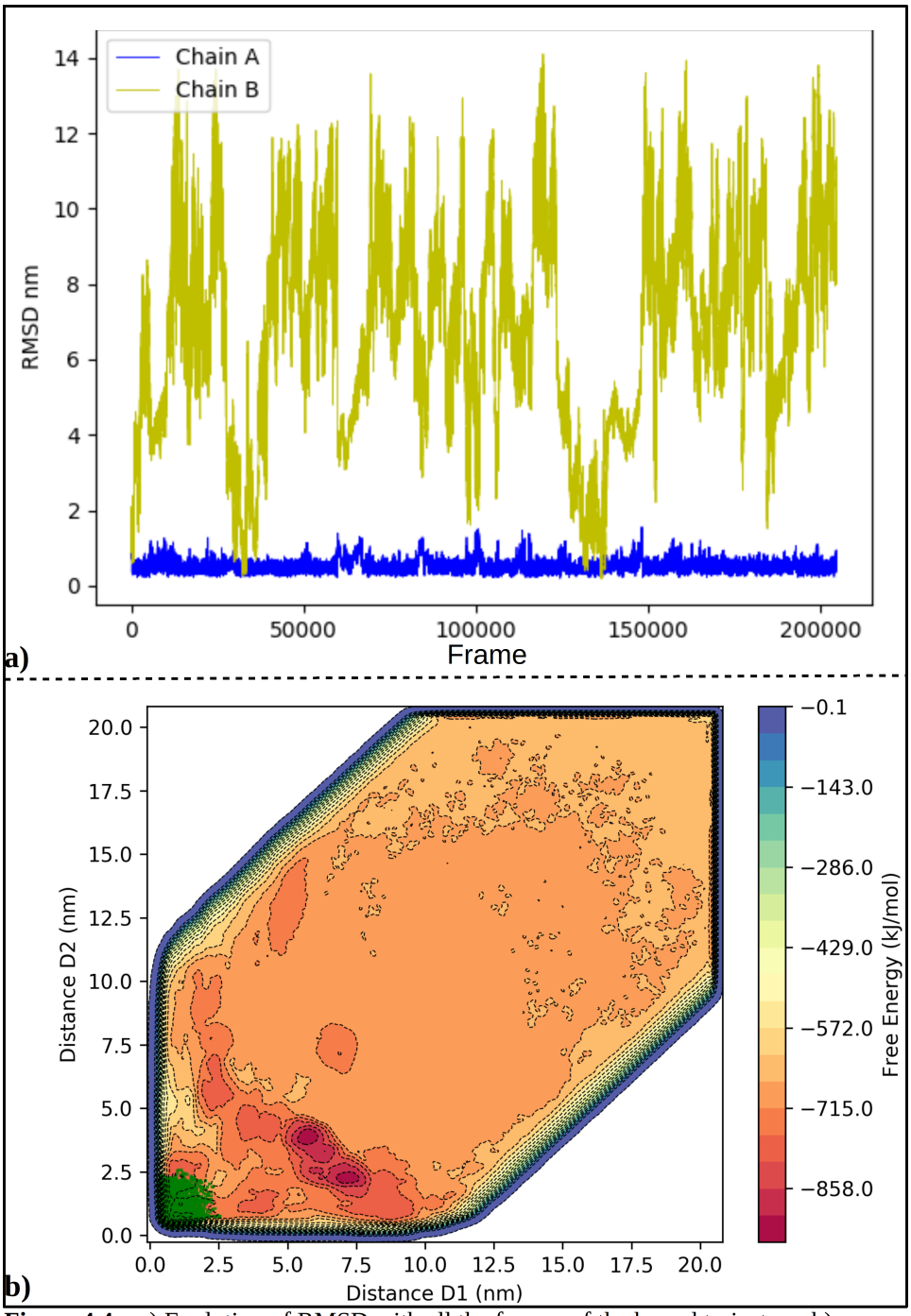


Figure 4.4 – a) Evolution of RMSD with all the frames of the bound trajectory. b) Projection of structures selected based on RMSD (green) on the FES.

Following the calculation of energies for bound state, the unbound state was selected as those configurations were chain A and chain B are not in contact, which would be all not included in

the bound trajectory. The result of the calculations is shown in Table-1 and the regions included in the unbound state are highlighted in green on the FES in Fig. 4.5.

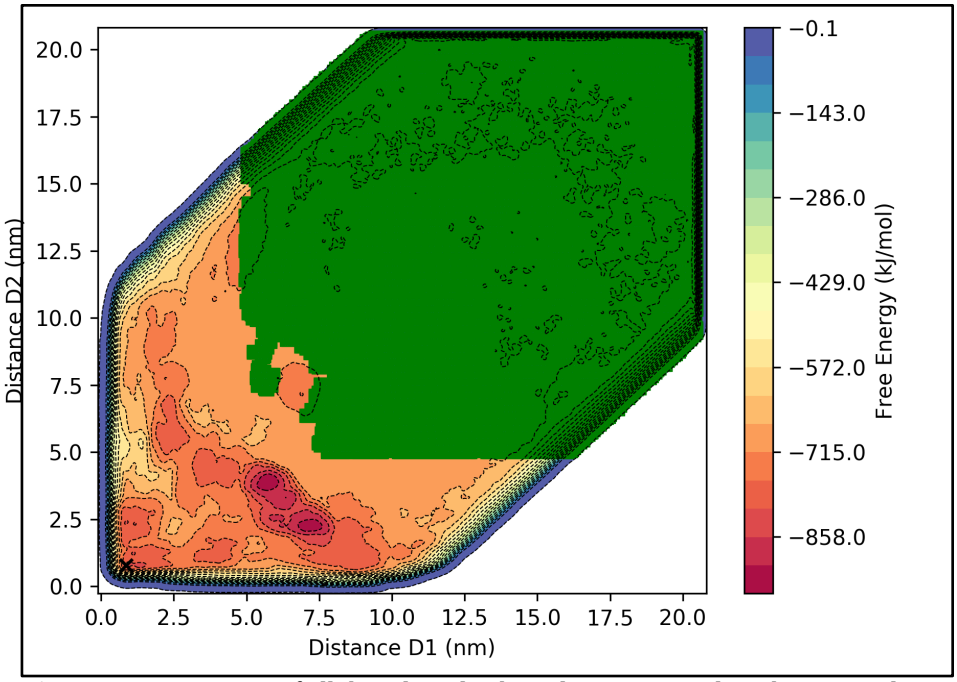


Figure 4.5 – Projection of all the selected unbound structures and CV bins around them (green).

4.2 Calculation of binding affinity

After obtaining the free energies for different states from metadynamics, the change in standard state Gibbs free energy and the corresponding binding affinities were calculated using the expressions described in section 3.5. The results are shown in Table-2. The bound states defined using criteria a) and c) are the most stable states with a nanomolar K_D. This result is in a similar range to the K_D of (0.41 ± 0.15) obtained for rat endophilin A1 by FRET analysis [24]. While the clustering was able to identify similar structures, it was not able effectively classify the most stable bound state and the calculated K_D was in the millimolar range.

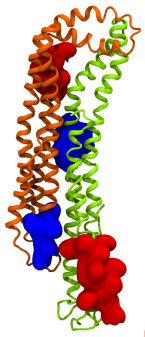
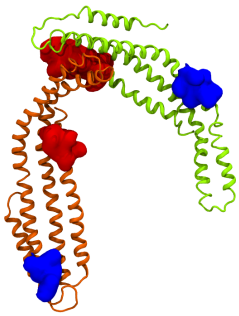
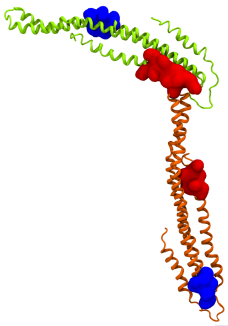
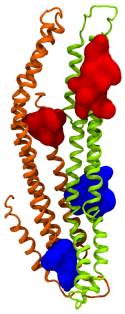
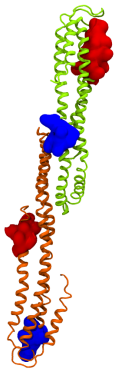
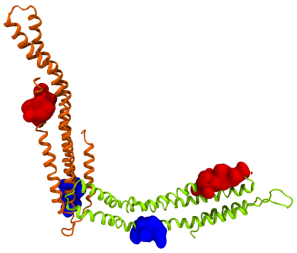
Table-2 Thermodynamic differences and binding affinities

State	$\Delta G_{meta} = G_{unbound} - G_{bound}$	$\Delta V_{P2} = V_{unbound} - V_{bound}$	Correction term	$\Delta G^{o}_{unbound-bound}$	K_D
Crystal Structure	72.271	0.854	23.13	49.190	5.14 nM
RMSD based	80.121	-0.146	23.13	56.982	0.25 nM
SC-1	34.371	-0.146	23.13	11.230	12.8 mM

4.3 Intermediate States

Intermediate states can be defined as configurations encountered by the system that are not part of the bound states. Since these states are stable, configurations belonging to these states will have higher population and can be identified as different clusters as shown in Fig. 4.2(a). To evaluate the thermodynamics of the the intermediates, CV bins around each cluster centroid were intergrated to calculate free energies. The energies were calculated for all the configurations corresponding to the selected CV bins for each cluster. For each cluster a representative structure closest to the centroid is also shown. It is important to keep in mind that this structure may not be an actual intermediate since the CV space is degenerate, but it can considered a close approximate. The structures and the thermodynamic quantities are displayed in Table-3.

Table-3 Structural representation of intermediates and their thermodynamic quantities

Cluster	Thermodynamics	Cluster	Thermodynamics
 C-0	ΔG_{meta} = 136.78 kJ/mol ΔU = -2.65 kJ/mol ΔS = -375.16 J/K-mol	 C-3	ΔG_{meta} = -16.29 kJ/mol ΔU = 55.82 kJ/mol ΔS = 307.28 J/K-mol
 C-1	ΔG_{meta} = -1.75 kJ/mol ΔU = 91.46 kJ/mol ΔS = 375.35 J/K-mol	 C-4	ΔG_{meta} = -10.5 kJ/mol ΔU = 278.91 kJ/mol ΔS = 1008.25 J/K-mol
 C-2	ΔG_{meta} = -49.49 kJ/mol ΔU = 110.68 kJ/mol ΔS = 591.35 J/K-mol	 C-5	ΔG_{meta} = -60.60 kJ/mol ΔU = 99.91 kJ/mol ΔS = 592.44 J/K-mol

Currently, at this stage of the simulation, the negative differences in free energies for clusters 1-5 are indicative of the fact that simulation has not yet converged.

4.4 Angle_CV simulation

Two simulations were performed with a different set of CVs and here I am presenting the one performed with a faster deposition time and higher height of the Gaussian (140 ps, 5 kJ/mol). This simulation, performed using a different set of CVs would ensure that the calculated free

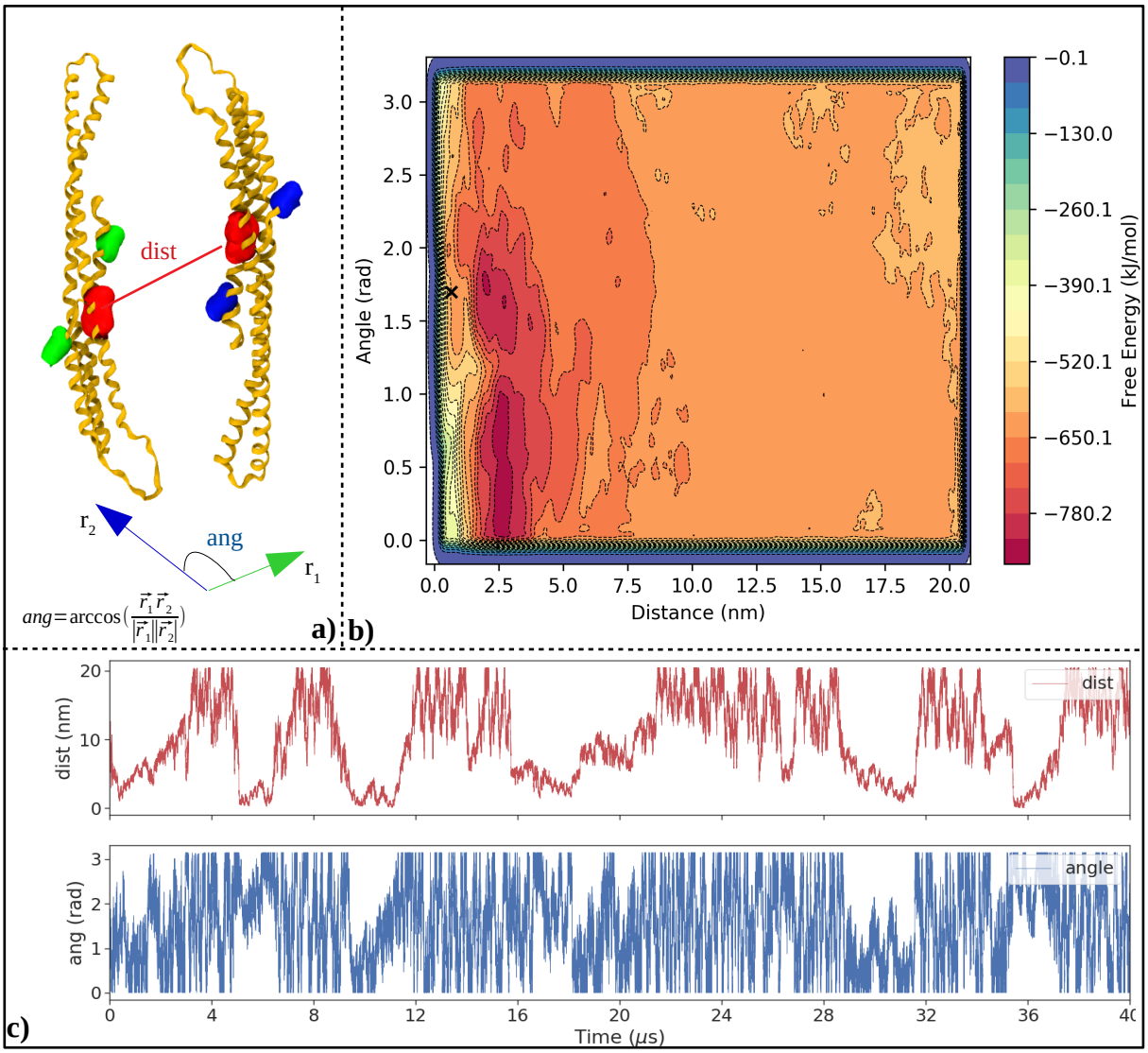


Figure 4.6 – a) Lsp1 chains A and B with the CVs, dist (red) and ang. The angle was calculated between r_1(green) and r_2 (blue). b) Contour plot of the free energy as a function of dist and ang after 40 µs. The '**X**' represents the CV values for the crystal structure. c) Progression of d1 (red) and d2 (blue) with time. The arrows represent transition events.

energy differences are robust. Here one CV was defined as the distance between two sites on chains A and B (different from the previous CVs) called ‘dist’. The second CV was defined as the angle between two vectors, where each vector joins the N terminal to the C terminal on each chain. These CVs are depicted in Fig. 4.6(a). The FES of the simulation, after 40 μs is shown in Fig. 4.6(b). The FES surface reveals two stable regions when the protein chains are close together at approximately 2.5 nm. These two regions correspond to the two chains being in different orientations. Interestingly, the protein chain prefers to be a bit further apart than their crystallized state. The evolution of the CVs in time is shown in Fig. 4.6(c). The protein chains start from far apart and explore different regions of the configurational space, driven by the bias potential. Over the trajectory they visit the stable minima regions multiple times but transition to the crystal structure orientation only once, at ~ 11 μs. To calculate the free energy differences, I followed a similar protocol to define the bound and the unbound states.

Angle Clustering

For this simulation, the clustering was performed similarly for only those structures where chains A and B are in contact. Here the features used for clustering were the CVs (ang,dist) with the number of clusters as 6. As seen from the clustering results in Fig. 4.7(a), the KMeans method is able to effectively classify regions only by distance and not by angle due to difference in the scales of the two features. Therefore in order to capture the bound state accurately, I performed another round of clustering for the C-1 cluster, but this time by normalizing the two features and classifying into 8 clusters. The results achieved are shown in Fig 4.7(b).

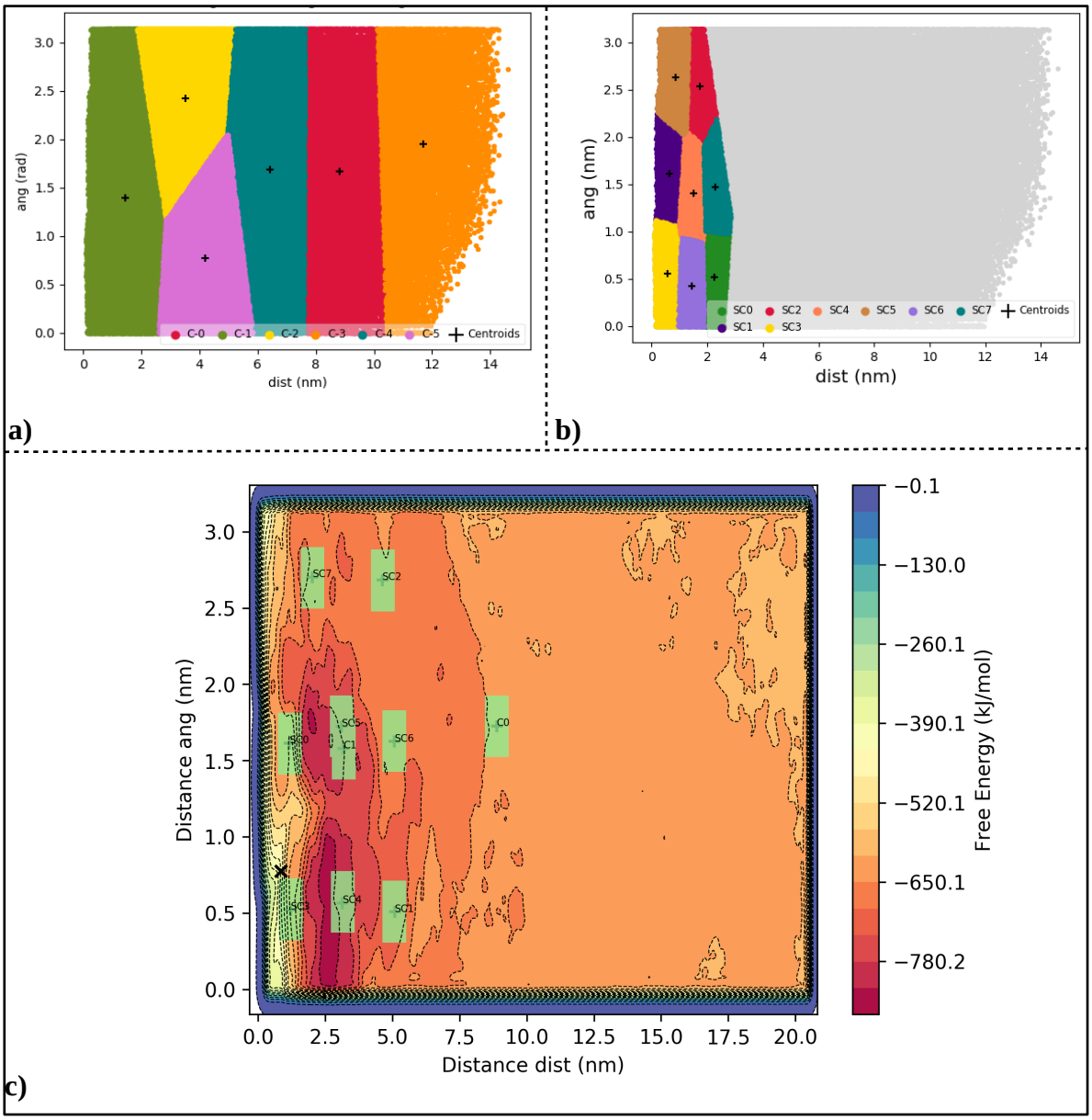


Figure 4.7 – a) Plot of 6 clusters as a function of dist,ang. b) Sub-cluster plot representing C-1 into 8 clusters. c) Projection of cluster centroids and CV bins around then on FES.

With this higher resolution of clusters, I identified two centroid regions, that correspond to the minima regions on the FES in Fig. 4.6(b), as the bound states. Similar to the previous simulation, the regions around the crystal structure were also picked as a possible bound state. The regions along with the cluster centroids are projected on the FES in Fig. 4.7(c). In addition to these

states, a fourth bound states was defined using an RMSD calculation. Similar to the protocol defined previously, the RMSD was calculated and is plotted w.r.t. the trajectory frames in Fig ().

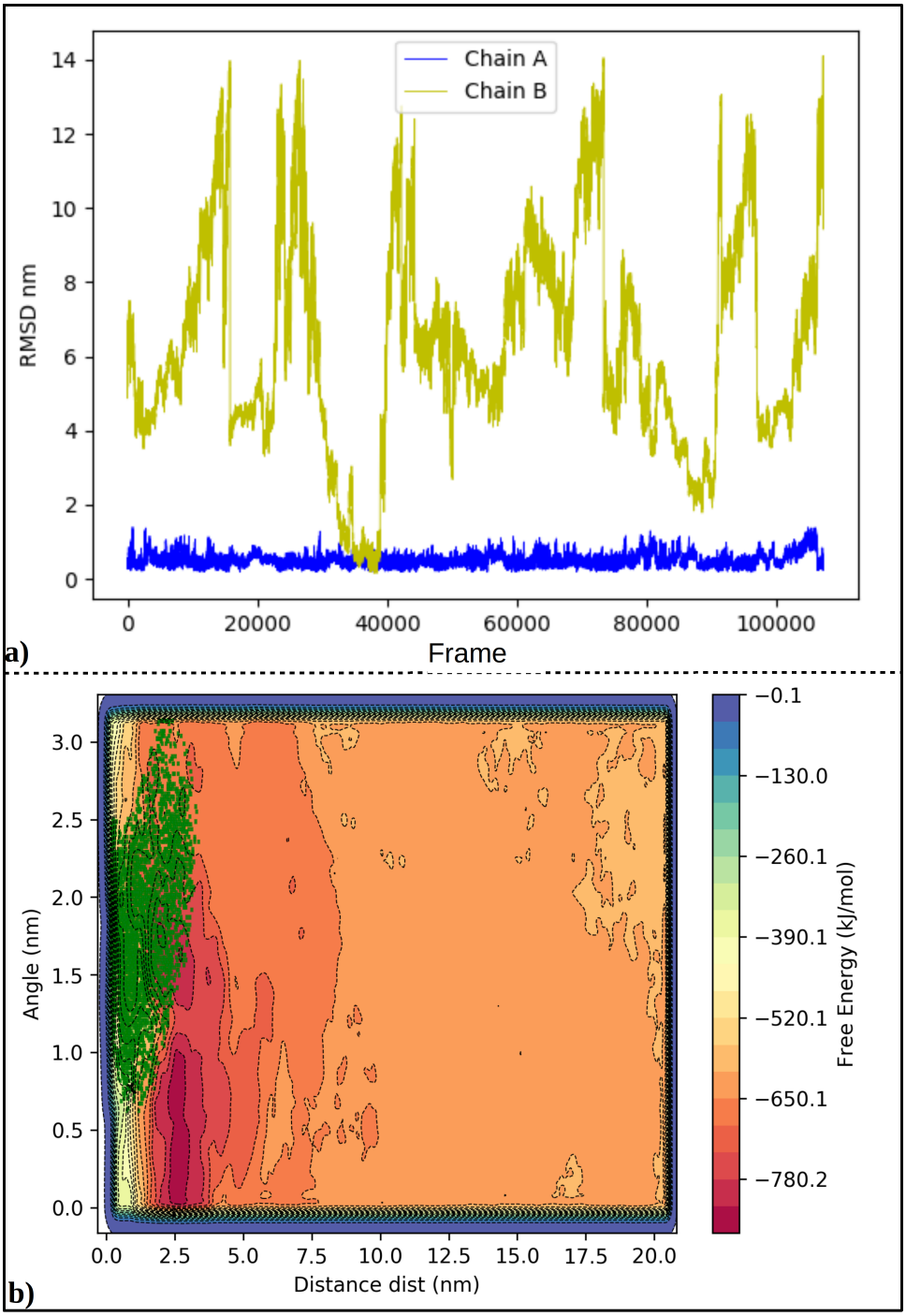


Figure 4.8 – a) Evolution of RMSD with all the frames of the bound trajectory. b) Projection of structures selected based on RMSD (green) on the FES.

As seen from the above plot, the system transitions into the crystal structure state once, which corresponds to the event observed at ~11μs in Fig. 4.8(a). The bound state was defined by considering only those structures with an RMSD < 1 nm. These structures are projected on the FES by their CV values and is shown in Fig. 4.8(b). With the above definitions of the bound states, the differences in thermodynamic quantities were calculated and are shown in Table-4

Table-4 Thermodynamic differences and binding affinities

State	**ΔG_{meta} = $G_{unbound}$ - G_{bound}**	**ΔV_{P2} = $V_{unbound}$ - V_{bound}**	**Correction term**	**$\Delta G^{o}_{unbound-bound}$**	**K_D**
Crystal Structure	-40.22	0.0	23.13	-63.354	
SC4	118.079	-2.0	23.13	94.825	~ 0.0001 pM
C1	85.046	-3.0		61.73	~ 1 pM
RMSD based	96.675	-1.013	23.13	73.48	~ 0.1 pM

The high free energy differences indicate that the simulation has not yet converged. Specifically, the system requires additional sampling in the crystal state. It is difficult to decide when to stop a metadynamics simulation and depends on the result being obtained. In addition for protein-protein systems it is also challenging to assess convergence. An indication of possible convergence would be when one observes a diffusive behavior of the biased CVs. This happens because the free energy surface has flattened and the motion in CV spaces is essentially a random walk. At this point in time, the bias potential oscillates around the true estimate of the free energy.

4.5 Structural Features

Besides evaluating the thermodynamic quantities of the dimerization process, I was also able to identify some structural trends across the simulations. One of them is related to the orientation of the protein chains in the dimer state. The crystal state of BAR dimer may not be the most stable state and the dimer prefers to be present in in different structural orientations. The FES in Fig. 4.6(b) appears to indicate this fact, as the stable regions are when the protein chains are further apart than the crystal state. To further investigate this, I performed a clustering analysis with different features for the two simulations. For the first one, the clustering was performed for the bound trajectory of the dist_CV simulation. The two features use were dist and RMSD. The dist variable was calculate in the same way as the one in the angle_CV simulation. These two features would provide an idea on how the orientation of the protein chains change with distance.

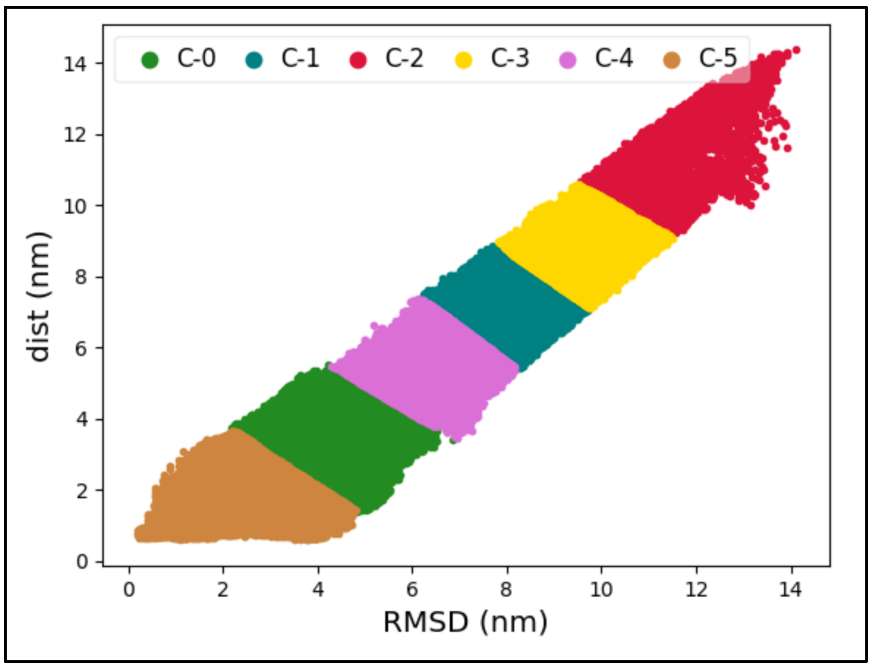


Figure 4.9 – Cluster representation based on dist and RMSD features. n=6 clusters.

The clustering result is shown in Fig. 4.9(a). The relationship of dist and RMSD follows the expected trend, that as the two chains move closer to each other they start to adopt the crystal structure increasingly. A closer look at the cluster C-5 reveals that at dist < 2 nm, the RMSD fluctuates between 0-5 nm.

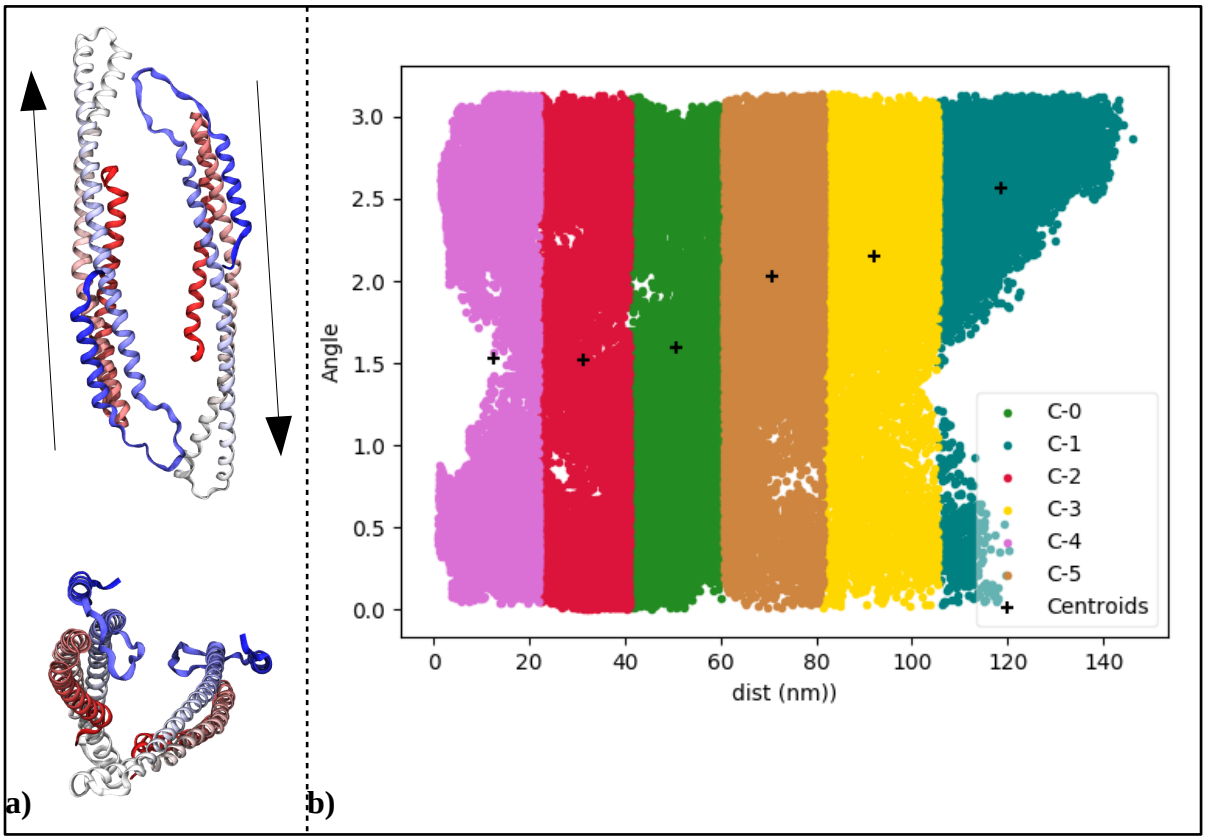


Figure 4.10 – a) Lsp1 chains colored from blue (N-terminus) to red (C-terminus). (top) The black arrows represent the axial vectors to calculate ax-ang. (bottom) Axial view of Lsp1. b) Cluster representation based on dist and ax-ang features. n=6 clusters.

A second clustering was performed for the bound trajectory of angle_CV simulation. Here the two feature selected were dist and ax-ang. The ax-ang is defined as the angle between the two vectors along the vertical axis of each chain as shown in Fig. 4.10(a). The clustering result is presented in Fig. 4.10(b). As seen from the plot, the ax-ang samples the complete range as the

protein chains move apart. However, when the protein chains are near, it prefers to be in two different orientations. One orientation range from ~ (2 – pi) radians and the second from ~ (0 to 1) radians. In the crystal structure, the ax-ang is 2.67 radians which would correspond to an anti-parallel orientation of the two chains, commonly observed for BAR proteins. Another interesting point here is that the dimer does not prefer to be between ~(1-2) radians which could indicate possible steric clashes. Together, these two results support the fact that the dimer structure is not strictly restricted to the crystal state orientation.

Chapter 5

Conclusion

This study has presented a comprehensive evaluation of the BAR protein dimerization process in terms of the thermodynamics and structural aspects. From the simulations, I have observed that binding affinity is sensitive to the choice of bound states and that defining these states may not be trivial. Nanomolar K_D was computed for certain states which is comparable to the subnanomolar K_D reported by experiments. Structural intermediates were also characterized that have highlighted the complex nature of the configurational space. An important caveat here is that the simulations have not yet converged which is also why the K_D calculated from the angle_CV simulation is very low. However, this has not prevented from observing certain structural features like the presence of different orientations of dimer and the rigidity of helical regions.

Importantly, this study also underlies the advantage of using MD simulations to characterize experimentally challenging processes. In addition the application of metadynamics can be used to overcome problems of timescales and sampling, which have historically impeded the use of MD simulations to investigate biological interactions.

Bibliography

1. Frost A, Unger VM, De Camilli P. The BAR Domain Superfamily: Membrane-Molding Macromolecules. *Cell*. 2009;137(2):191-196. doi:10.1016/J.CELL.2009.04.010

2. Itoh T, De Camilli P. BAR, F-BAR (EFC) and ENTH/ANTH domains in the regulation of membrane–cytosol interfaces and membrane curvature. *Biochim Biophys Acta - Mol Cell Biol Lipids*. 2006;1761(8):897-912. doi:10.1016/J.BBALIP.2006.06.015

3. Peter BJ, Kent HM, Mills IG, et al. *BAR Domains as Sensors of Membrane Curvature: The Amphiphysin BAR Structure*. Vol 303.; 2004. https://www-jstor-org.proxy1.library.jhu.edu/stable/pdf/3836080.pdf?refreqid=excelsior%3A8cfd9d1895368464aa07d93ddcfcc084. Accessed September 10, 2019.

4. Tarricone C, Xiao B, Justin N, et al. *The Structural Basis of Arfaptin-Mediated Cross-Talk between Rac and Arf Signalling Pathways*. Vol 411.; 2001. www.nature.com. Accessed September 11, 2019.

5. Masuda M, Mochizuki N. Structural characteristics of BAR domain superfamily to sculpt the membrane. *Semin Cell Dev Biol*. 2010;21(4):391-398. doi:10.1016/J.SEMCDB.2010.01.010

6. Rahman A, Rank F, Stillincer H. *Molecular Dynamics Study of Liquid Water**. Vol 55.; 1971.

7. Rahman A. *Correlations in the Motion of Atoms in Liquid Argon**.

8. Mccammon JA, Gel BR, Karp Ius M. *Dynamics of Folded Proteins*. Vol 267. Nature Publishing Group; 1977.

9. Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE. Biomolecular Simulation: A Computational Microscope for Molecular Biology. 2012. doi:10.1146/annurev-biophys-042910-155245

10. Leach AR. Molecular modelling: principles and applications. Second. *Pearson Educ Ltd Essex Engl*. 2001;410:457-460.

11. Lim BBC, Lee EH, Sotomayor M, Schulten K. Molecular Basis of Fibrin Clot Elasticity. *Structure*. 2008;16(3):449-459. doi:10.1016/j.str.2007.12.019

12. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K. Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics. *Structure*. 2008;16(5):673-683. doi:10.1016/j.str.2008.03.005

13. Chandler D. Introduction to modern statistical. *Mech Oxford Univ Press Oxford, UK*. 1987.

14. *Molecular Dynamics Simulations of Biomolecules*.; 2002. http://www.nature.com/naturestructuralbiology. Accessed April 12, 2020.

15. Lee EH, Hsin J, Sotomayor M, Comellas G, Schulten K. Discovery Through the Computational Microscope. *Structure*. 2009;17(10):1295-1306. doi:10.1016/j.str.2009.09.001

16. Laio A, Parrinello M. Escaping free-energy minima. *Proc Natl Acad Sci U S A*. 2002;99(20):12562-12566. doi:10.1073/pnas.202427399

17. Barducci A, Bonomi M, Parrinello M. Metadynamics. *Wiley Interdiscip Rev Comput Mol Sci*. 2011;1(5):826-843. doi:10.1002/wcms.31

18. Bussi G, Branduardi D. Free-Energy Calculations with Metadynamics: Theory and Practice. In: ; 2015:1-49. doi:10.1002/9781118889886.ch1

19. Valsson O, Tiwary P, Parrinello M. Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint MD: molecular dynamics. 2016. doi:10.1146/annurev-physchem-040215-112229

20. Mcdonald NA, Gould KL. Cell Cycle Linking up at the BAR: Oligomerization and F-BAR protein function Linking up at the BAR: Oligomerization and F-BAR protein function. 2016. doi:10.1080/15384101.2016.1190893

21. Dislich B, Than ME, Lichtenthaler SF. Specific amino acids in the BAR domain allow homodimerization and prevent heterodimerization of sorting nexin 33. *Biochem J*. 2011;433(1):75-83. doi:10.1042/BJ20100709

22. Gortat A, Jouve San-Roman M, Vannier C, Schmidt AA. Single Point Mutation in Bin/Amphiphysin/Rvs (BAR) Sequence of Endophilin Impairs Dimerization, Membrane Shaping, and Src Homology 3 Domain-mediated Partnership * □ S. 2011. doi:10.1074/jbc.M111.325837

23. Jennifer L Gallop1,3,4, Christine C Jao2 3, Helen M Kent1, P Jonathan G Butler1, Philip R Evans1,*, Ralf Langen2 *, and Harvey T McMahon1 *. Mechanism of endophilin N-

BAR domain mediated membrane curvature. *EMBO J*. 2006. https://www.embopress.org/doi/pdf/10.1038/sj.emboj.7601174. Accessed September 11, 2019.

24. Capraro BR, Shi Z, Wu T, et al. Kinetics of Endophilin N-BAR Domain Dimerization and Membrane Interactions *. 2013. doi:10.1074/jbc.M112.435511

25. Karotki L, Huiskonen JT, Stefan CJ, et al. Eisosome proteins assemble into a membrane scaffold. *J Cell Biol*. 2011;195(5):889-902. doi:10.1083/jcb.201104040

26. Arkhipov A, Yin Y, Schulten K. Membrane-Bending Mechanism of Amphiphysin N-BAR Domains. *Biophys J*. 2009;97(10):2727-2735. doi:10.1016/J.BPJ.2009.08.051

27. Yu H, Schulten K. Membrane Sculpting by F-BAR Domains Studied by Molecular Dynamics Simulations. *PLoS Comput Biol*. 2013;9(1):1002892. doi:10.1371/journal.pcbi.1002892

28. Blood PD, Swenson RD, Voth GA. Factors Influencing Local Membrane Curvature Induction by N-BAR Domains as Revealed by Molecular Dynamics Simulations. *Biophys J*. 2008;95(4):1866-1876. doi:10.1529/BIOPHYSJ.107.121160

29. Arkhipov A, Yin Y, Schulten K. Four-Scale Description of Membrane Sculpting by BAR Domains. *Biophys J*. 95:2806-2821. doi:10.1529/biophysj.108.132563

30. Blood PD, Voth GA. Direct observation of Bin/amphiphysin/Rvs (BAR) domain-induced membrane curvature by means of molecular dynamics simulations. *Proc Natl Acad Sci*. 2006;103(41):15068-15072. doi:10.1073/PNAS.0603917103

31. Monticelli L, Kandasamy SK, Periole X, Larson RG, Peter D, Marrink S-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. doi:10.1021/ct700324x

32. Deng Y, Roux BB. Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant. 2006. doi:10.1021/ct060037v

33. Boresch S, Tettinger F, Leitgeb M, Karplus M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. 2003. doi:10.1021/jp0217839

34. Doudou S, Burton NA, Henchman RH. Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force. doi:10.1021/ct8002354

35. Hansen N, Van Gunsteren WF. Practical Aspects of Free-Energy Calculations: A Review. 2014. doi:10.1021/ct500161f

36. Michel J, Essex JW. Prediction of protein-ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. doi:10.1007/s10822-010-9363-3

37. Perthold JW, Oostenbrink C. 5697−5708 Cite This. *J Chem Theory Comput*. 2017;13. doi:10.1021/acs.jctc.7b00706

38. Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J Comput Phys*. 1977;23(2):187-199. doi:10.1016/0021-9991(77)90121-8

39. Saglam AS, Chong LT. Protein-protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations †. 2019. doi:10.1039/c8sc04811h

40. Ceccarelli M, Anedda R, Casu M, Ruggerone P. CO escape from myoglobin with metadynamics simulations. *Proteins Struct Funct Bioinforma*. 2007;71(3):1231-1236. doi:10.1002/prot.21817

41. Gervasio FL, Laio A, Parrinello M. Flexible Docking in Solution Using Metadynamics. 2005. doi:10.1021/ja0445950

42. Barducci A, Bussi G, Parrinello M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. 2008. doi:10.1103/PhysRevLett.100.020603

43. Bussi G, Gervasio FL, Laio A, Parrinello M. Free-Energy Landscape for Hairpin Folding from Combined Parallel Tempering and Metadynamics. 2006. doi:10.1021/ja062463w

44. Raiteri P, Laio A, Gervasio FL, Micheletti C, Parrinello M. Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics †. 2006. doi:10.1021/jp054359r

45. Ziółkowska NE, Karotki L, Rehman M, Huiskonen JT, Walther TC. Eisosome-driven plasma membrane organization is mediated by BAR domains. 2011;8(5). doi:10.1038/nsmb.2080

46. Periole X, Cavalli M, Marrink S-J, Ceruso MA. Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. doi:10.1021/ct9002114

47. Periole X;, Cavalli M;, Marrink S-J;, Ceruso MA. Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *J Chem Theory Comput*. 2009;5(9):2531-2543. doi:10.1021/ct9002114

48. Pronk S, Páll S, Schulz R, et al. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*. 2013;29(7):845-854. doi:10.1093/bioinformatics/btt055

49. Yesylevskyy SO, Schäfer L V., Sengupta D, Marrink SJ. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. Levitt M, ed. *PLoS Comput Biol*. 2010;6(6):e1000810. doi:10.1371/journal.pcbi.1000810

50. Marrink SJ, Tieleman DP. Perspective on the Martini model. *Chem Soc Rev*. 2013;42(16):6801. doi:10.1039/c3cs60093a

51. Michalowsky J, Schäfer L V, Holm C, Smiatek J. Cite as. *J Chem Phys*. 2017;146:163319. doi:10.1063/1.4974833

52. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: New feathers for an old bird. *Comput Phys Commun*. 2014;185(2):604-613. doi:10.1016/j.cpc.2013.09.018

53. van Rossum G. Python tutorial, May 1995. *CWI Rep CS-R9526*. 1995;(CS-R9526):1-65. http://oai.cwi.nl/oai/asset/5007/05007D.pdf.

54. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J Comput Chem*. 2011;32(10):2319-2327. doi:10.1002/jcc.21787

55. Gowers RJ, Linke M, Barnoud J, et al. *MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations*.; 2016. http://mdanalysis.org. Accessed May 2, 2020.

56. Pedregosa FABIANPEDREGOSA F, Michel V, Grisel OLIVIERGRISEL O, et al. *Scikit-Learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot*. Vol 12.; 2011. http://scikit-learn.sourceforge.net. Accessed May 2, 2020.

57. Rotkiewicz P, Skolnick J. Fast procedure for reconstruction of full-atom protein models from reduced representations. *J Comput Chem*. 2008;29(9):1460-1465. doi:10.1002/jcc.20906

58. Humphrey W, Dalke A, Schulten K. {VMD} -- {V}isual {M}olecular {D}ynamics. *J Mol Graph*. 1996;14:33-38.

59. Stone J. \em An Efficient Library for Parallel Ray Tracing and Animation. April 1998.

60. Sanner M, Olsen A, Spehner J-C. Fast and Robust Computation of Molecular Surfaces. In: *Proceedings of the 11th ACM Symposium on Computational Geometry*. New York: ACM; 1995:C6-C7.

61. Gilson MK, Given JA, Bush BL, McCammon JA. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys J*. 1997;72(3):1047-1069. doi:10.1016/S0006-3495(97)78756-3

62. General IJ. A Note on the Standard State's Binding Free Energy. doi:10.1021/ct100255z

63. Gilson MK, Given JA, Bush BL, McCammon JA. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys J*. 1997;72(3):1047-1069. doi:10.1016/S0006-3495(97)78756-3

64. Lloyd SP. Least Squares Quantization in PCM. *IEEE Trans Inf Theory*. 1982;28(2):129-137. doi:10.1109/TIT.1982.1056489

65. Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. *Appl Stat*. 1979;28(1):100. doi:10.2307/2346830

Curriculum Vitae

Education

AUGUST 2018 – MAY 2020
M.S.E Chemical and Biomolecular Engineering
JOHNS HOPKINS UNIVERSITY, BALTIMORE
CGPA – 4.0/4.0
Courses – Thermodynamics and Statistical Mechanics, Molecular Immuno-engineering, Modeling in the Living Cell, Kinetic Processes, Non-Equilibrium Physics of Cells and Molecules

JULY 2014 – JUNE 2016
M.Tech Bioprocess Technology
INSTITUTE OF CHEMICAL TECHNOLOGY, MUMBAI
CGPA – 8.75/10.0
Courses – Bioreaction Engineering, Biosystems engineering, Unit operations in Bioprocess Lab – Chemical engineering, Microbiology, Fermentation and downstream processing

JULY 2010 – MAY 2014
B.Pharm (Hons.)
BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI
CGPA – 7.87/10.0
Courses – General Biology, Thermodynamics, Microbiology, Pharmacology, Biological Chemistry

Research experience

JANUARY 2019 - PRESENT
M.S.E Research, **JOHNSON LAB, DEPT. OF BIOPHYSICS, JOHNS HOPKINS UNIVERSITY**
Molecular dynamics simulations of Lsp-1 (yeast BAR domain protein) to study dimerization on membrane surface. Application of metadynamics to calculate free energy of dimerization on membrane surface and in solution. **Principal Investigator:** Dr. Margaret Johnson

DECEMBER 2017 – MAY 2018
Research trainee, **CENTRE FOR CELLULAR AND MOLECULAR BIOLOGY, HYDERABAD, INDIA**
Part of the bioinformatics team in analyzing Mnase seq data to develop nucleosome occupancy profile. Involved with several bioinformatics tools and programming in Python.

JUNE 2015 – JUNE 2016
M.Tech Research project, **INSTITUTE OF CHEMICAL TECHNOLOGY, MUMBAI, INDIA**
Project involved improving lipid production by microalgae through media optimization. Developed a scaled up system for algal cultivation. **Research Supervisor:** Dr. J.T. Waghmare

JULY 2013 – DECEMBER 2013
Intern, LAB OF MOLECULAR ONCOLOGY, CDFD, HYDERABAD, INDIA
Project was to identify possible sites of gene fusion in Colorectal Cancer. aCGH data combined with qPCR was used to narrow down sites of recombination most likely to harbor fusions.

MAY 2012 – JULY 2012
Summer intern, ORCHID BIOMED, GOA, INDIA
Studied different functions of the Quality Control department during the manufacturing of a product. Worked in the R&D lab with the different RDT and ELISA products.

Work Experience

JULY 2016 – JUNE 2017
Management trainee, CIPLA BIOTEC, GOA, INDIA
Worked as part of the production team. Responsibilities included media preparation, bioreactor operation and monitoring, cell culture sampling and other upstream activities. Operated single use bioreactors up to a scale of 2000L and mixing systems. Performed scale-up and cell culture calculations (IVCC, productivity). Also part of the Quality management team.

Skills

- MD simulations: GROMACS, LAMMPS, MDAnalysis.
- Programming languages – Java, Python, MATLAB
- Bioinformatics – Bowtie, samtools
- Aseptic cell culture handling
- Biological techniques – Gel electrophoresis, PCR, q-PCR, GC-MS, UV-spectrometer

Activities

- **Publications:**

  Critical review titled “Lactoferrin: health benefits, technology and applications” published in Agro FOOD Industry Hi Tech Vol 26(6) 2015. Journal Impact factor 0.225

- **Academic Projects:**
  - Study Oriented Project : – ‘Role of Telomerase Inhibitors in Cancer therapy’
  - Android Application ‘GluCheck’- Android App to monitor a person’s glucose level.
  - GIS based X-ray imaging – Use of a software and GIS imaging to improve X-ray analysis.
  - Website designing for LAMP course: Designed a customized website for BITS - Medical Centre as part of the course project.
- Member of National Service Scheme (2010-2012)
- Involved in several NSS programs like Schooling of rural children, Tree plantation and Blood Donation Camp.
- Hobbies- Reading Books, magazines, Newspaper, Listening to music
- Fan of Arsenal Football Club